Current State and Capture Plan for EDRN Data

Heather Kincaid

- Data Capture Plan
  - specific applications
  - types of data
  - schedule for data capture
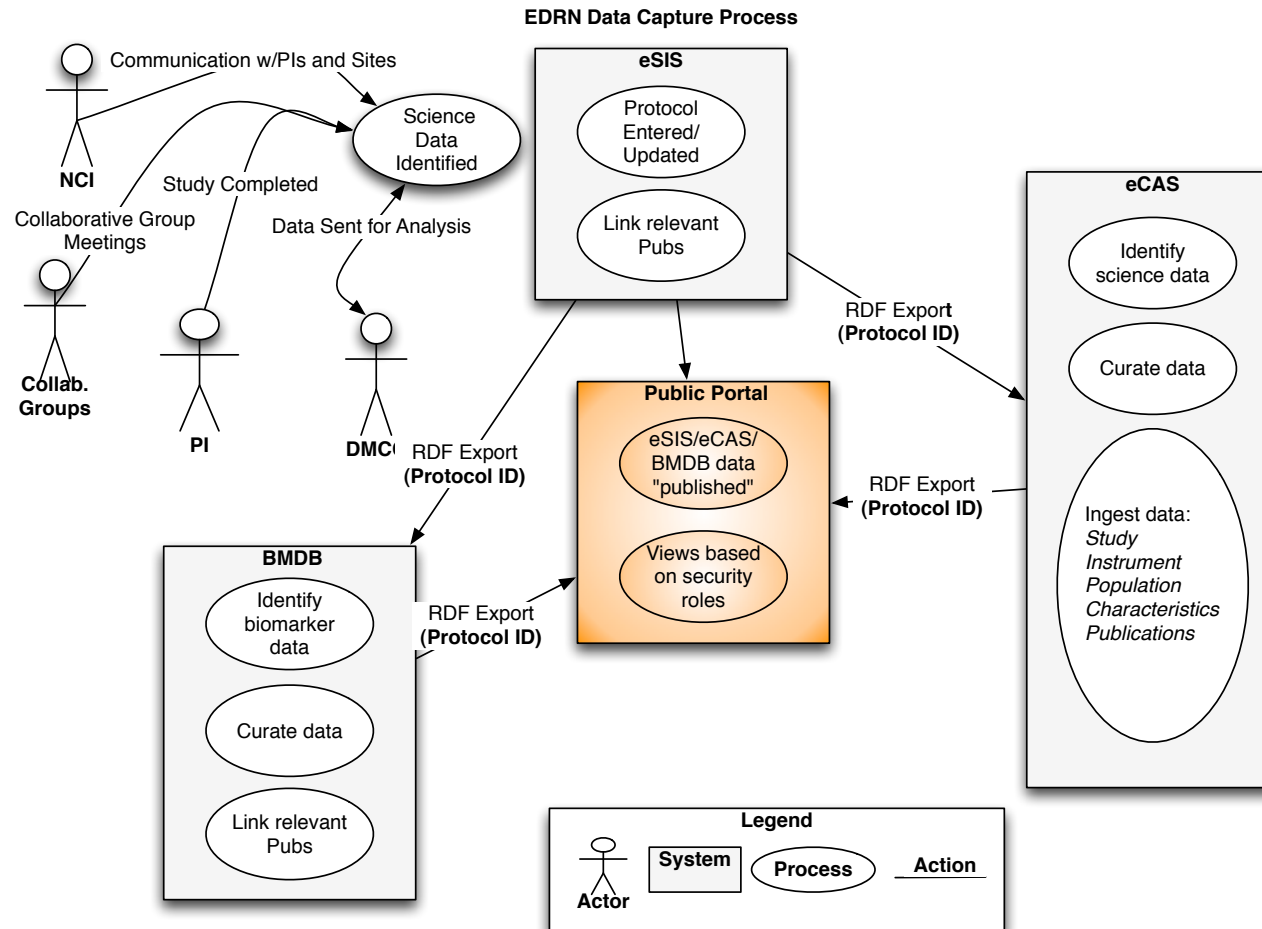- Access to data
- Security

Types of Data

- Biomarkers
- Research Data
- Publications
- Instrument Data
- Other Related Data – KEGG, etc.
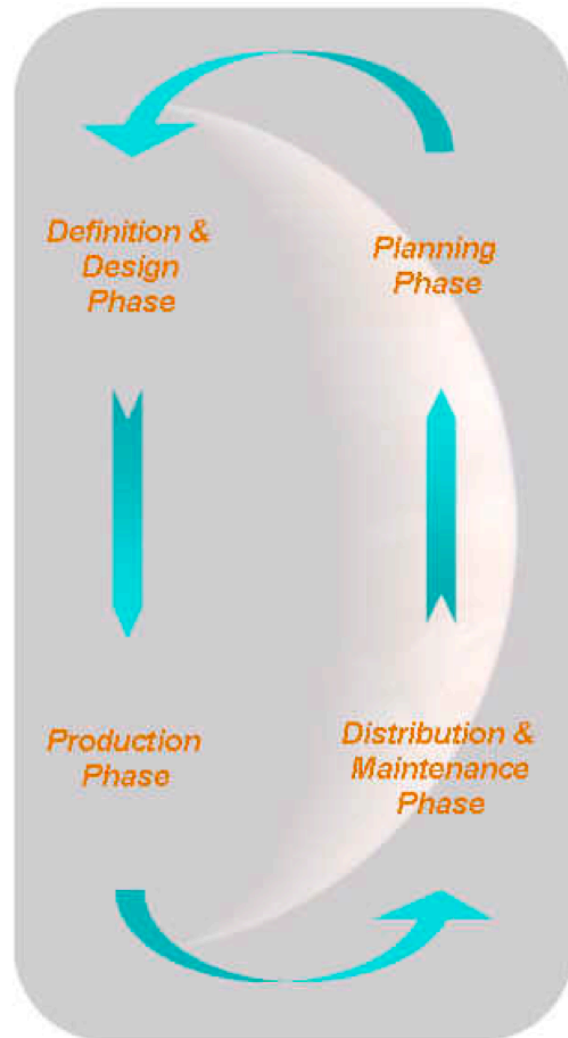
Applications

- Biomarker Database
  - Biomarker Metadata
  - Links to
    - Protocol (Study) Information
    - Publications
    - Related Data (web resources such as KEGG)
- eCAS
  - Research Data (Metadata and "unprocessed" and "frozen" datasets)
  - Instrument Data
  - Links to
    - Protocol (Study) Information
    - Publications
- eSIS
  - Protocol Information
- ERNE
  - Specimens
- Atlas

- Other Data?

EDRN Data Capture Process

EDRN Data Capture Process

NCI — Communication w/PIs and Sites

Collaborative Group Meetings

Study Completed

Data Sent for Analysis

Collab. Groups

PI

DMC

Science Data Identified

eSIS
- Protocol Entered/Updated
- Link relevant Pubs

RDF Export **(Protocol ID)**

eCAS
- Identify science data
- Curate data
- Ingest data: *Study Instrument Population Characteristics Publications*

Public Portal
- eSIS/eCAS/BMDB data "published"
- Views based on security roles

RDF Export **(Protocol ID)**

RDF Export **(Protocol ID)**

BMDB
- Identify biomarker data
- Curate data
- Link relevant Pubs

RDF Export **(Protocol ID)**

Legend
Actor | System | Process | Action

Acquisition Schedule

Application	Short Term (next 6 months)	Long Term (next 5 years)
BMDB	•Data that's in eCAS (if the marker is far enough along) •What/Who determines what is curated? 　•SELDI 　•WHI 　•Grizzle Markers	Deep Content Data analyzed at DMCC Meetings (SC, Scientific, etc.) Colab. Group Calls Data Sharing Plans and Renewals
eCAS	•List provided by DMCC Statisticians •DABS Datasets •Data related to markers curated in the BMDB (DCP, CA125, etc.)	Data analyzed at DMCC Meetings (SC, Scientific, etc.) Colab. Group Calls Data Sharing Plans and Renewals
eSIS	Populated by Curator when contacting investigators for BMDB or eCAS data	System developed b/w DMCC and NCI to notify DMCC when new Protocol s/b entered.
ERNE	Reference Sets	Add sites with large specimen repositories.
Atlas (Lung, Colon)		

Planning Phase:
- Data archiving requirements written into mission Announcement of Opportunity
- Pre-proposal briefing on PDS data archiving requirements given to potential proposers
- Proposal data archiving section reviewed by PDS
- PDS orientation to flight project staff
- Data archiving working groups formed

Definition & Design Phase:
- Project Data Management and Archive Plans define data to be archived
- Data Product and Volume Organization Software Interface Specifications detail the data and volume structure
- Preliminary metadata labels loaded into PDS catalog

Production Phase:
- Raw and processed data products, labels (metadata) and documentation produced
- Preliminary and quick-look data made accessible via Project and PDS web pages
- Data archive products validated and peer-reviewed; liens corrected

Distribution & Maintenance Phase:
- Final data products made available on-line
- PDS add the data to the archive
- Physical copies sent to NSSDC
- PDS provides data, documentation and science expertise to users
- Data archive maintained via periodic media refreshes, addition of new / updated data products

EDRN Data Lifecycle

- How and at what point to we capture research data?

- What is a research data timeline and at what stages do we capture data.

- What happens at each phase?

- How do we relieve burden for sites?

Security

- How do we address security?